基于分类外形搜索的人脸特征点定位 *

黄玉琴，潘华伟†

(湖南大学 信息科学与工程学院，长沙 410082)

摘　要：针对传统由粗糙到精准的人脸外形搜索方法，其每一次外形搜索需要在整个外形搜索空间进行，提出一种基于分类的外形搜索方法。该方法始于一个包含不同人脸形状的外形搜索空间，首先利用基于相关性的特征选择方法对随机森林分类器进行优化，利用训练的随机森林分类器将外形搜索空间分为若干个外形搜索子空间；然后根据输入样本和随机森林分类器确定与当前外形最接近的外形搜索子空间，并计算对应子空间的中心和对应样本的后验概率分布，方便后续阶段更好地进行外形搜索；最后采用级联回归进行人脸特征点定位。在 300-W 数据集上的实验结果表明，此方法不仅有效降低了外形搜索的时间，同时在无约束环境中具有良好的鲁棒性。

关键词：人脸外形搜索；随机森林；级联回归；人脸特征点定位；由粗到精

中图分类号：TP391.41　　doi: 10.3969/j.issn.1001-3695.2017.11.0774

Face alignment based on classified shape searching

Huang Yuqin[1], Pan Huawei[2†]

(*School of information Science & Engineering Hunan University*, *Changsha* 410082, *China*)

Abstract: Aiming at improving the traditional shape searching which needs to search in the whole shape space each time. A new approach is proposed based on classified face shape searching. This approach begins with a shape space that contains diverse shapes, first we optimize the random forest classifiers by the feature selection method of correlation-based and train the random forest classifiers by training samples, then divides the shape into several sub-spaces by random forest classifiers, and search the sub-space that most similar to the current shape, then estimate the center of the sub-space and probability distribution. Finally, employs cascaded regression to realize face alignment. The approach demonstrates its obvious decreases in searching time and its good robustness in unconstrained environment on the 300-W database.

Key Words: face shape searching; random forest; classified regression; face alignment; course-to-fine

0　引言

人脸特征点定位是计算机视觉中的一个基本问题，是人脸识别[1~3]、人脸表情分析[4~6]和 3D 人脸重建[7,8]中的关键技术。准确高效的人脸特征点定位为上述问题奠定了良好的基础。人脸特征点定位技术旨在预测给定 2D 人脸图像特征点的位置。在实际应用中，由于受到人脸姿态、表情变化、光照变化以及遮挡等因素的影响，无约束条件下的人脸特征点定位一直是研究难点与热点。

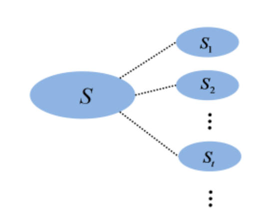
人脸特征点定位是一个非凸优化问题，现有的很多方法分两步解决该问题，分别为人脸样本初始化和迭代优化人脸外形。由于是非凸问题，样本的初始化很大程度上会影响特征点定位精度，所以一个好的初始化外形是精确定位的基础。文献[9,10]利用平均外形作为初始外形。这种初始化的方法在人脸检测效果不准确和人脸偏转角度较大的时候效果很差，容易陷入局部最优。文献[11,12]采用多重初始化的方法，该方法在进行外形回归的过程中以不同的初始值多次执行算法得到多个回归结果，并取多次回归结果的中位数作为最终结果，这种方法缺乏与真实样本是否匹配的依据，初始化次数需要交叉验证来设置，实验结果并不理想且缺乏泛化能力。Zhu 等人[13]提出一种由粗糙到精准的外形搜索方法，该方法开始于一个粗糙的外形搜索空间，每次迭代回归之前根据当前人脸形状以及概率分布在整个外形搜索空间里进行一次外形搜索，得到与样本最为接近的外形搜索子区域，然后根据子区域内外形以及对应的后验概率求子区域中心并利用该值更新原样本空间的概率分布，为后续外形搜索做准备，这样反复迭代直至收敛。这种初始化方法提供了一个比较准确的初始外形，但是由于每次外形搜索都要对整个样本空间进行，耗时较长，效率比较低。

针对以上问题，本文提出一种基于分类的外形搜索方法。该方法与级联回归交替使用。传统的外形搜索方法每一次外形搜

基金项目：广东省科技计划资助项目（2013B090600021）；国家科技部资助项目（2014BAK08B01）

作者简介：黄玉琴（1991-），女，硕士研究生，主要研究方向为机器学习、图像处理；潘华伟，男（通信作者），助理教授，博士，主要研究方向为计算机图形学、虚拟现实（38697014@qq.com）.

索都需要在整个外形搜索空间进行，需要根据上一阶段搜索得到的外形子空间中心来更新整个样本空间的概率分布，计算量很大。为了提高实验效率，本文首先采用基于相关性的特征选择方法对随机森林分类器进行优化，利用训练样本训练随机森林分类器并对训练样本进行分类，之后每一次外形搜索都在分类得到的外形搜索子空间进行，该方法在很大程度上减少了计算量。实验结果表明，该方法的效率有明显提高，同时在 *300-W* 数据集上具有良好的鲁棒性。本文方法的外形子空间搜索过程如图 1 所示。其中(a)中 S 是整个外形搜索空间，S_k 是第 k 个外型搜索子空间;(b)中黄色点表示子空间中心，红色点表示真实形状。

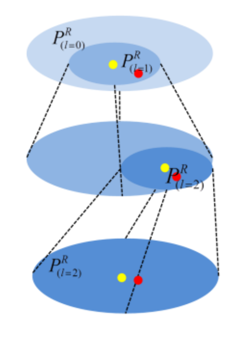


(a)数据集分类

Stage1: given $P^R_{(l=0)}$,estimate $\bar{S}_{(l=1)}$,estimate $P^R_{(l=1)}$

Stage2: given $P^R_{(l=1)}$, estimate $\bar{S}_{(l=2)}$, estimate $P^R_{(l=2)}$

Stage3: given $P^R_{(l=2)}$, estimate $\bar{S}_{(l=3)}$



(b)在子空间进行的外形搜索

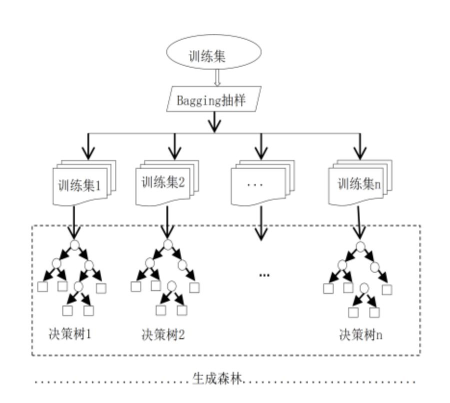
图 1　本文方法的外形子空间搜索过程

1　随机森林简介

随机森林，是用随机的方式建立一个森林，指的是利用多棵决策树对样本进行训练并预测的一种分类器。简单来说，随机森林就是由多棵决策树构成的，森林中每棵树之间是没有关联的。对于每棵树，使用的训练集都是从总的训练集中采用有放回方法抽取得到的，这意味着，总的训练集中有些样本可能多次出现在一棵树的训练集中，也可能一次都不会出现。

得到随机森林后，当有新样本输入时，就让森林中每棵树做一次判断，看看该样本属于哪一类，最后的分类结果通过森林中所有决策树的投票结果进行决策。

随机森林是在 Bagging[14]的基础上发展的，其随机性体现在两个方面，首先是每棵决策树的训练样本都是通过在原训练集中采用有放回抽取得到的，其次构建决策树分裂节点时可供选择的属性集也是随机选取的。由于这两个随机属性的结合，降低了森林中随机树之间的相似度，提升了随机森林的分类精度。随机森林算法如图 2 所示。



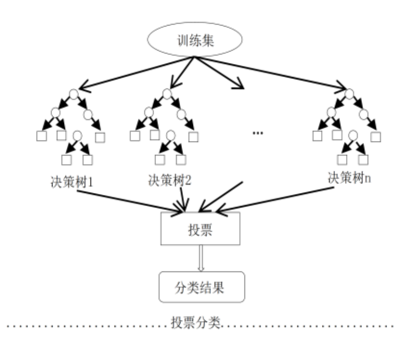


图 2　随机森林算法示意图

2　基于分类的外形搜索算法

随机森林分类方法具有高效性和不会过拟合等优点，为了更好地对训练样本进行分类，缩小外形搜索空间，提高外形搜索效率，本文引入基于相关性的特征选择方法对随机森林分类器进行优化。即在特征池中采用基于相关性的特征选取方法选取若干个特征组成新的特征集作为单棵决策树的输入，使得决策树的分类性能更好。采用改进的随机森林分类器实现对整个外形搜索空间的分类，将样本空间 S 分为 2^F （本文 $F=5$ ）类[15]，表示为 $S=\{S_1,\ldots,S_{2^F}\}$ 。每一类表示一个外形搜索子空间，分类目标是使得每一个外形搜索子空间中包含的训练样本在姿态、表情、光照和遮挡等方面基本一致。

2.1　基于相关性的特征选择

对于所有训练样本，选取 P 个相同位置上的像素点，分别计算两个像素之间的差值，可以得到 n (P^2) 个形状索引特征（shape-indexed features）。为了提高决策树的分类精度， 通常采用的方法是在一个随机生成的特征池里面选出最具有代表性的特征，这种方法叫做 *n-Best*[16,17]，n 表示特征池的大小。由表 1 可知，随着特征数量的不断增加，误差变化不明显，但训练时间大大增加。为了解决该问题，本文采用基于相关性的特征选择方法选取特征，即利用特征之间的相关性来进行特征选择。

基于相关性特征选择的关键步骤是计算特征之间的相关性，然后根据相关性强弱来选择特征。本文采用基于互信息[18]的方法来度量特征之间的相关性。对于一个特征 α_i ，用 $P(\alpha_i)$ 表示特征 α 取第 i 个值的概率，$P(\alpha_i\,|\,\beta_j)$ 表示特征 β 取 β_j 时 α 取 α_i 的概率。α 的信息熵 $H(\alpha)$ 和已知 β 时 α 的条件信息熵 $H(\alpha\,|\,\beta)$ 计算方法如下：

$$H(\alpha) = -\sum_i P(\alpha_i) lb P(\beta_j) \quad (1)$$

$$H(\alpha \mid \beta) = -\sum_j P(\beta_j) \sum_i P(\alpha_i \mid \beta_j) lb P(\alpha_i \mid \beta_j) \quad (2)$$

式（2）中 lb 表示以 2 为底的对数。特征 α ， β 之间的互信息 $M(\alpha,\beta)$ 计算方法如下：

$$\begin{aligned} M(\alpha,\beta) &= H(\alpha) - H(\alpha \mid \beta) = H(\beta) - H(\beta \mid \alpha) \\ &= \sum_{\alpha,\beta} p(\alpha\beta) lb \frac{p(\alpha\beta)}{p(\alpha)p(\beta)} \end{aligned} \quad (3)$$

特征 α 和 β 之间的相关性大小采用式（4）来度量。

$$S(\alpha,\beta) = 2\left[\frac{M(\alpha,\beta)}{H(\alpha)+H(\beta)}\right] \quad (4)$$

相关性 $S(\alpha,\beta)$ 的取值范围为 $[0,1]$ ， $S(\alpha,\beta)=0$ 表示两个特征不相关， $S(\alpha,\beta)=1$ 表示两个特征完全相关。

计算两两特征之间的相关性之后，利用特征聚类方法[19]根据特征之间的相关性强弱对所有特征进行聚类（下文表示为 t 类），同一类中的特征具有较强的相关性，不同类中的特征相关性弱。聚类之后从每一类中随机选取一个特征组成新的特征集作为决策树的输入。

2.2 改进的随机森林分类器

决策树是一种树型分类器，由根节点，中间节点和叶子节点组合而成，是随机森林的组成单元。由于人脸样本会受到表情、姿态、光照、遮挡等外在条件的影响，为了保证单棵决策树分类的精度，本文引入基于相关性的特征选择方法，根据特征之间的相关性聚类后选择 t 个特征作为单棵决策树的输入（见 2.1 节），从而提高随机森林分类器的分类精度和鲁棒性。

随机森林由若干树型分类器 $\{h(x,\theta_k), k=1,2,\ldots\}$ 随机组合而成，其中 x 是输入， $\{\theta_k\}$ 是独立同分布的随机变量，当有新样本输入时，最终的分类结果由所有单棵决策树的分类结果投票决定，具体步骤如下：

a)利用 Bagging 方法在样本空间 S 中的 N 个样本中有放回的抽取 N 个样本作为个别训练集 S_N ；

b)对于每一个个别训练集 S_N ，生成一棵决策树（见算法 1）；

c)重复 a)b)，直到生成 m 棵随机树；

d)对未知类别的样本进行分类时，输出的类别标签由森林中树的多数投票决定，即

$$c = \arg\max_c \left(\frac{1}{m}\sum_{k=1}^{m} I(h(x,\theta_k)=c)\right) \quad (5)$$

随机森林在 Bagging 的基础上引入随机选择的属性，更大程度上降低了森林中树之间的相关性，保证了随机森林的分类性能。

算法 1 构建单棵决策树

输入：个别训练集 S_N (样本 $s_i \Leftrightarrow (\alpha_1,\alpha_2,\ldots,\alpha_n)$ α_i 为样本的特征)的 n 个特征中采用 2.1 提出的基于相关性特征选择方法抽取出的 $t(t \ll n)$ 个特征，在生成整个森林过程中 t 保持不变。

输出：单棵决策树。

a)首先判别当前节点中所有样本是否属于同一类别。如果属于同一类别，判定当前节点为叶子节点，样本标签即为叶子节点类别；否则执行步骤 b)；

b)判别当前节点中所有样本是否为空。如果为空集，判定当前节点为叶子节点，样本标签即为叶子节点类别；否则执行步骤 c)；

c)利用基于相关性的特征选取方法（见 2.1 节）从所有特征中选取 t 个特征，分别计算每个特征 α_i 对于当前节点训练集的信息增益[20]，选择信息增益最大的特征作为分裂节点进行分裂。得到新的分支后，返回步骤 a)；

d)当所有分支都充分生长、所有样本均达到叶子节点，随机树建立完成。

2.3 分类外形搜索算法

在文献[13]提出的由粗到精的(Course-to-Fine)外形搜索方法的基础上首先利用改进的随机森林分类器将样本空间 S 分为 2^F （本文 $F=5$ ）类，分类过程如图 1(a)，每一类表示一个外形搜索子空间。输入新的样本时，根据输入图像特征和分类器得到与之最接近的一个外形搜索子空间。之后的外形所搜在该外形搜索子空间进行（外形搜索过程如图 1(b)）。相对于传统的由粗到精的外形搜索方法每一次的外形搜索都在整个外形搜索空间进行，本文方法减少了每一次外形搜索过程中的计算量，提高了实验效率。

3 实验分析

3.1 分类特征的选择

本文引入基于相关性的特征选择方法，选择具有识别力的特征进行随机森林分类器的训练。对所有训练样本，选取 P 个相同位置的像素点，分别计算两个像素之间的差值，可以得到 n （ P^2 ）个形状索引特征（shape-indexed features），通过基于相关性的特征选择方法选取相关性较高的 t 个特征作为单棵决策树的输入。表 1 列出了 *n-Best* 方法与本文方法在特征选择上对误差以及时间的影响。

表 1 *n-Best* 与本文方法的比较

	1-Best	64-Best	1024-Best	本文方法
误差	5.61	5.13	4.98	**3.11**
时间(s)	0.2	2.9	110.5	**0.13**

3.2 分类精度验证

为了比较本文引入基于相关性特征的选择方法对随机森林进行优化后随机森林的分类性能，这里对单棵决策树、随机森林分类器（Forest_RC）以及优化后的随机森林分类器在不同 UCI 数据集[21]上的分类误差进行了比较分析，分析结果如表 2 所示。

表 2 不同数据集的分类误差比较

UCI 数据集	C4.5	Forest_RC	优化的森林
Glass	0.5013	0.4831	0.3962
Heart	0.3091	0.2957	0.2304
Pima	0.3265	0.2711	0.2530
Balance	0.4019	0.3851	0.3042
Vehicle	0.3990	0.3224	0.2980

在本实验中随机森林中树的数量会影响最终的分类结果，进行了多次交叉实验，由图 3 可以看出树的数量到 100 之后结果收敛，为了控制计算量，本文把树的数量设为 100。

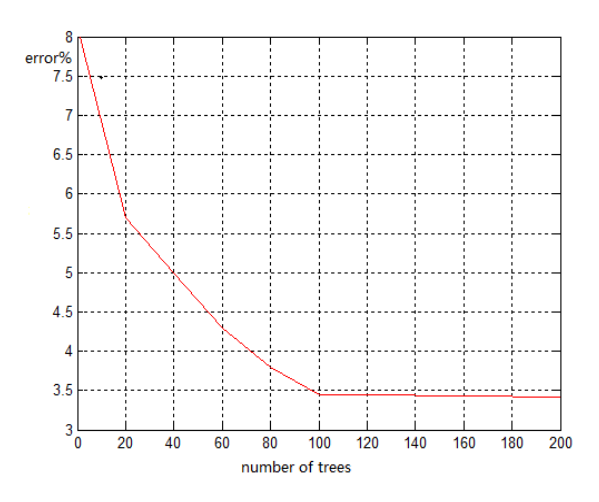


图 3　随机森林中树的数量和误差的关系

3.3　人脸特征点定位的实验分析

3.3.1 数据集和参数设置

本文实验在 *300-W*（300 Face in-the-Wild）数据集[22]上进行。该数据集是在已有数据集的基础上建立的，这其中包含标准化的带 68 个标记点的数据集 AFW，LFPW，HELEN 以及 XM2VT5，同时还包括一个包含 135 张图像的具有挑战性的 IBUG 数据集。把数据集 AFW，LFPW，HELEN 中的所有训练样本（一共 3 148 张图像）作为训练集，另外，做两个部分的测试（一共 689 张图像）：把数据集 LFPW，HELEN 中的测试样本作为普通测试集；把 IBUG 数据集中的 135 张图像作为另一组测试集。

综上，组成一个带有 68 个标记点的 3148 张训练图像和 689 张测试图像的 *300-W* 数据集。实验环境：CPU 频为 2.6Hz，4G 内存，500G 硬盘，Windows7 64 位操作系统，仿真软件 MATLAB 2012a。

为了权衡实验精度和速度，本文通过交叉验证把单棵树的深度 F 设置为 5，实验表明增加树的深度只会改善边缘误差。搜索阶段 L 的值设为 3。本文采用的 *300-W* 数据集在聚类时 t 的取值为 23。本文设置其他参数如下：S=3148 （训练样本），$m=100$，P=400，$n=500$，$L=3$，$K_l=3$，$N_l=10$，$T=10$，P=400。由于本文搜索算法有不错的容错率，在搜索时为了追求较高的运行效率，在级联回归阶段采用复合特征，即在前两级（$l=1,2$）采用 BRIEF 特征[23]，在最后一级采用 SIFT 特征。

3.3.2 实验结果分析

从表 3 可以看到，由于利用改进的随机森林做分类，每一次的外形搜索只需要在外形搜索子空间进行，待测样本只需要与对应的外形搜索子空间中的样本进行后验概率计算，极大减少了运算量，因此本文的改进外形搜索算法相对于传统方法在运行效率上有了明显提高。

从表 4 的实验结果可以看出，本文的算法在准确性预测上要明显优于以往传统的级联回归方法。但由于 CFSS 三级都采用 SIFT 特征，而本文方法采用了复合特征，即在前两个阶段采用了比较简单的 BRIEF 特征，在最后一阶段采用 SIFT 特征，因而在准确性上相对于以前的方法改善效果相对比较小，但值得指出的是在准确预测相差不多的情况下，本文的方法的运行速度要优于其他方法。

表 3　各方法运行效率比较

方法	fps
RCPR	26
SDM[24]	28
DRMF[25]	34
CFAN[7]	35
CFSS[13]	40
本文	58

表 4　各方法预测平均误差

方法	普通测试集	IBUG 测试集
RCPR	6.18	7.01
SDM	5.57	6.32
DRMF	6.65	6.93
CFAN	5.50	6.75
CFSS	4.73	5.83
本文	4.63	5.42

为了更直观的看出本文方法的定位精度，图 4 给出了本文采用的方法与几种主要方法在 *300-W* 数据集上的累计误差分布图（cumulative errors distribution，CED）。

图 4 中的 x 坐标表示人脸特征点定位归一化平均误差，y 坐标表示测试图像归一化平均误差低于 x 值的对应百分比。从图中曲线可以看出，相比于其他的算法，本文算法即使在具挑战性的数据集上还是能够取得比较好的实验结果。

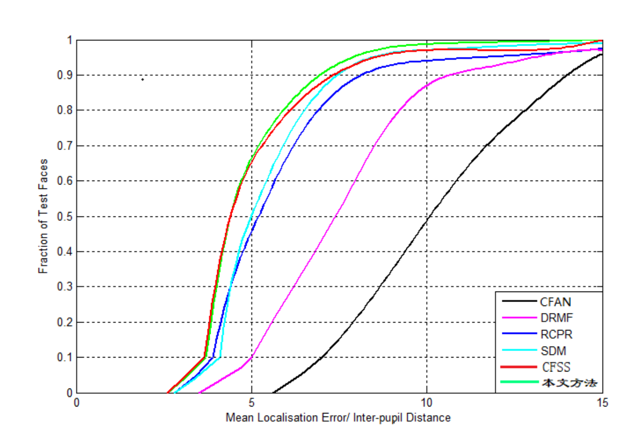
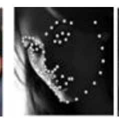
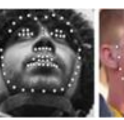


图 4　300-W 数据集上的 CED 图

图 5 从上到下给出了本文方法以及传统外形搜索方法在 *300-W* 数据集上的部分实验结果。从实验结果可以看出，本文方法在定位精度上相比于之前的方法明显改进。但是值得指出的是，在定位精度相当的情况下，本文的定位效率有大幅度提高。



(a)本文方法下的结果



(b)CFSS 方法下的结果



(c)SDM 方法下的结果

图 5 不同方法部分实验结果比较

4 结束语

为了提高传统的外形搜索特征点定位方法的运行效率，本文提出一种基于分类的外形搜索方法。该方法首先引入基于相关性的特征选择方法对随机森林分类器进行改进，利用改进的随机森林分类器对整个人脸样本空间进行分类，然后在对应的样本子空间中进行外形搜索，得到最接近于当前样本的人脸外形，最后采用级联回归来增强预测的精度。由于分类需要对大量特征进行相关性计算，处理大数据集的时候运算量比较大，会影响运行效率，下一步本文将试图探索其他的方法来进行样本分类并在运行效率和定位精度上做进一步的改进。

参考文献：

[1] Lawrence S, Giles C L, Tsoi A C, et al. Face recognition: a convolutional neural-network approach [J]. IEEE Trans on Neural Networks, 1997, 8 (1): 98-113.

[2] Oneto L, Ghio A, Ridella S, et al. Out-of-sample error estimation: the blessing of high dimensionality [C]// Proc of IEEE International Conference on Data Mining Workshop. 2015: 637-644.

[3] Burgos-Artizzu X P, Perona P. Robust face landmark estimation under occlusion [C]// Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2013: 1513-1520.

[4] Kaburlasos V G, Papadakis S E, Papakostas G A. Lattice computing extension of the FAM neural classifier for human facial expression recognition [J]. IEEE Trans on Neural Networks & Learning Systems, 2013, 24 (10): 1526-1538.

[5] Weng Y, Cao C, Hou Q, et al. Real-time facial animation on mobile devices [J]. Graphical Models, 2014, 76 (3): 172-179.

[6] Bradley D, Bradley D, Beeler T, et al. Real-time high-fidelity facial performance capture [J]. ACM Trans on Graphics, 2015, 34 (4): 46.

[7] Zhang J, Shan M. et al. Course-to-fine auto-encoder networks (CFAN) for real-time face alignment [C]// Lecture Notes in Computer Science, vol 8690. 2014: 1-16.

[8] Li Q, Cheng G, Fan J, et al. Embracing the blessing of dimensionality in factor models [J]. Journal of the American Statistical Association, 2016.

[9] Asthana A, Zafeiriou S, Cheng S, et al. Robust discriminative response map fitting with constrained local models [C]// Computer Vision and Pattern Recognition. 2013: 3444-3451.

[10] Tzimiropoulos G, Pantic M. Optimization problems for fast AAM fitting in-the-wild [C]// Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2013: 593-600.

[11] Burgosartizzu X P, Perona P, Dollar P. Robust face landmark estimation under occlusion [C]// Proc of IEEE International Conference on Computer Vision. 2014: 1513-1520.

[12] Sun J, Wen F, Wei Y, et al. Face alignment by Explicit Shape Regression [J]. International Journal of Computer Vision, 2012, 107 (2): 177-190.

[13] Zhu S, Li C, Chen C L, et al. Face alignment by coarse-to-fine shape searching [C]// Computer Vision and Pattern Recognition. 2015: 4998-5006.

[14] Pavlidis T. IEEE transactions on pattern analysis and machine intelligence-special memorial issue for professor King-Sun Fu [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2004, 26 (5): 0_1-673.

[15] ÃZuysal M, Calonder M, Lepetit V, et al. Fast key point recognition using random ferns [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2010: 448-461.

[16] Ozuysal M, Calonder M, Lepetit V, et al. Fast keypoint recognition using random ferns [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2010: 448-461.

[17] Liu R, Shen J, Sun Q, et al. Cascaded pose regression revisited: face alignment in videos [C]// Proc of the 3rd IEEE International Conference on Multimedia Big Data. 2017: 291-298.

[18] Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy [J]. Journal of Machine Learning Research, 2004: 1205-1224.

[19] 蒋盛益，王连喜. 基于相关性的特征选择 [J]. 计算机工程与应用, 2010, 46 (20): 153-156.

[20] Dosse M B, Kiers H A L, Berge J M F T. Anisotropic generalized Procrustes analysis [J]. Computational Statistics & Data Analysis, 2011, 55 (5): 1961-1968.

[21] University of California-Irvine. UCI repository of machine learning database [EB/OL]. http://www. ics. edu/~mlearn/MLRepository. html.

[22] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 faces in-the-wild challenge: the first facial landmark localization challenge [C]// Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2013: 397-403.

[23] Ren S, Cao X, Wei Y, et al. Face alignment at 3000 FPS via regressing local binary features [J]. IEEE Trans on Image Processing, 2014, 25 (3): 1685-1692.

[24] Xiong X, Torre F D L. Supervised descent method and its applications to face alignment [C]// Computer Vision and Pattern Recognition. 2013: 532-539.

[25] Asthana A, Zafeiriou S, Cheng S, et al. Robust discriminative response map fitting with constrained local models [C]// Computer Vision and Pattern Recognition. 2013: 3444-3451.